-2-

## AMENDMENT TO THE CLAIMS

1. (Currently Amended)    A method of compressing a log of natural language data, the log having a plurality of natural language help query strings relative to a help function of a computer system, each help query string including at least two words~~tokens~~, the method comprising:

    applying a subsumption~~compression~~ operation to each help query string, wherein the subsumption operation identifies a single word difference between the help query string and another help query string~~each string is a query relative to a help function of a computer system~~;

    ~~identifying two strings that match each other after the compression operation; and~~

    removing one of the two help query strings~~matching~~ strings from the log to form a compressed log; and

    ~~feeding the compressed log back to the compression operation at least once; and~~

    training a statistical process with the compressed log.

2. (Previously Presented)    The method of claim 1, wherein the log is a log of user-initiated inputs to a help interface.

3-7. (Canceled)

8. (Currently Amended)    The method of claim 1~~7~~, wherein subsumption includes applying an impossibility condition to selectively compute edit distance.

9. (Currently Amended)    The method of claim 1, and further comprising:

    applying a second subsumption~~compression~~ operation to each help query string;

    determining if any two strings match each other after the second subsumption~~compression~~ operation; and

    removing one of the two help query~~matching~~ strings from the log.

10-13. (Canceled)

14. (Currently Amended)     A system for compressing a query log having a plurality of natural language help query strings, each string having a plurality of <u>words</u><s>tokens</s>, the system comprising:

        an input for receiving a raw query log of natural language help query strings relative to a help function of a computer;

        memory for storing the raw query log;

        a processor for applying at least one <u>subsumption</u><s>compression</s> operation <u>wherein the subsumption operation identifies a single word difference between a help query</u> <s>to each</s> string <u>and another help query string</u>, and for <u>removing one of the help query strings</u><s>scanning the modified strings to determine if any match each other so that one of the matching strings can be removed</s> to form a compressed query log; and

        wherein the processor is configured to <s>feed the compressed query log back to the compression operation at least once and to</s> utilize the compressed query log to train a statistical process once the removal is complete.

15. (Canceled)

16. (Currently Amended)     The system of claim <u>14</u><s>15</s>, wherein each help-related query is relative to a computer system.

17-19. (Canceled)

20. (Original) The system of claim 19, wherein subsumption includes applying an impossibility condition to selectively compute edit distance.

21. (Currently Amended)     The system of claim 14, and further comprising:

      applying at least a second subsumption~~compression~~ operation to each string;

      determining if any two strings match each other after the second subsumption~~compression~~ operation; and

      removing one of the two matching strings from the log.


22-25. (Canceled)


26. (New)     The method of claim 1, wherein the method further comprises discarding the additional word and collapsing the pair of help query strings if the additional word does not significantly change the meaning.


27. (New)     The method of claim 26, wherein the subsumption operation includes a statistical operation relative to the additional word.


28. (New)     The method of claim 1, wherein the subsumption operation is absolute between an N word help query string and an (N-1) word help query string.


29. (New)     The method of claim 1, wherein the subsumption operation is guided by vocabulary features.


30. (New)     The method of claim 1, wherein subsumption is blocked if the additional word is in a control vocabulary.


31. (New)     The system of claim 14, wherein the system discards the additional word and collapses the pair of help query strings if the additional word does not significantly change the meaning.

32. (New)     The system of claim 31, wherein the subsumption operation includes a statistical operation relative to the additional word.

33. (New)     The system of claim 14, wherein the subsumption operation is absolute between an N word help query string and an (N-1) word help query string.

34. (New)     The system of claim 14, wherein the subsumption operation is guided by vocabulary features.

35. (New)     The system of claim 14, wherein subsumption is blocked if the additional word is in a control vocabulary.